

Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2017

1. Objectives

We report national-level prevalence estimates of chronic HBV derived by a systematic review of peer-reviewed literature reporting HBV prevalence (hepatitis B surface antigen [HBsAg]) in the general population for children under 5 years of age and population above 5 years of age.

The estimates are reported for 190 countries and territories organized into 6 WHO regions and by income status organized into 4 levels according to the World Bank classification.

The estimates are provided before vaccine introduction and after vaccine introduction (2015). We also estimate the number of people living with chronic HBV infection on a national, regional, and global level and address changes over time.

2. List the funding sources for the work.

This research was partially supported by the WHO's Initiative for Vaccine Research and the WHO Hepatitis Programme.

3. Data inputs

We undertook and report our updated systematic review in line with the criteria outlined in the PRISMA guidelines (Moher, et al., 2009). We updated the systematic review by (Schweitzer, et al., 2015) which included a systematic search on articles published between Jan 1, 1965, and Oct 23, 2013.

We updated the systematic search on articles published between Oct 23, 2013, and March 20, 2017 in the databases Embase, PubMed, Global Index Medicus, Popline, and Web of Science.

We developed a search strategy and adapted it for each database using a combination of Medical Subject Headings (MeSH) and free text including terms related to HBV and to prevalence (see appendix 1). We supplemented database searches by inspecting publications referenced in studies identified in the systematic search. In addition, we contacted the 194 WHO member states and requested them to review the outcome of our systematic search and provide published and unpublished data. We received feedback from the Member States until 20 March 2017.

Figure 1 shows the PRISMA flowchart diagram of the study selection process. Two authors (M Hasso and R Vargas) systematically screened search results and independently reviewed retrieved records applying the eligibility criteria. One author (DeLa Hoz) resolved queries and the inconsistencies. Two authors (A Vicari, and X Riveros) independently conducted quality assurance checks on 100% of newly identified reports.

For the sensitivity analyses, we assessed the representativeness of included study data and assigned the category “non-representative” to studies done in indigenous populations or in locations within countries known as particularly low or high endemicity areas for HBV. We assessed representativeness on the basis of information available from source manuscripts and author expert opinion for 90% of the reports identified. NB 100% will be done before peer review publication.

Data extraction

Following full text review, we extracted data from each study using the following variables: study characteristics (study and sample collection dates, study locations i.e., city, subnational [an area, region, state, or province in a country], or national level), participant characteristics (age range, sex, year, and population group), and prevalence of the HBV marker, type of laboratory tests, and number of participants the HBV marker prevalence was based on.

4. Specify the inclusion and exclusion criteria. Identify all ad-hoc exclusions.

The criteria were similar to (Schweitzer, et al., 2015). Observational studies on chronic HBV infection seroprevalence (HBsAg prevalence), done in the general population or among blood donors, health-care workers (HCWs), and pregnant women were considered for inclusion in this systematic review. Studies were excluded if they were systematic reviews or meta-analyses, surveillance reports, case studies, letters or correspondence, or did not contain HBsAg seroprevalence data. Studies were also excluded if they exclusively reported prevalence estimates for high-risk population groups (e.g., migrants and refugees). Eligible literature, identified in title or abstracts screening, was obtained for full text screening and grouped by country and WHO region. During full text screening, we excluded articles reporting data without specifying the serological marker.

5. Provide information on all included data sources and their main characteristics. For each data source used, report reference information or contact name/institution, population represented, data collection method, year(s) of data collection, sex and age range, diagnostic criteria or measurement method, and sample size, as relevant.

See above for systematic search details.

Following full text review, we extracted data from each study using the following variables: study characteristics (study and sample collection dates, study locations i.e., city, subnational [an area, region, state, or province in a country], or national level), participant characteristics (age range, sex, year, and population group), and prevalence of the HBV marker, type of laboratory tests, and number of participants the HBV marker prevalence was based on.

Data of eligible articles were entered into a Microsoft EXCEL[®] and/or Distiller databank by two reviewers independently. Information was extracted for author name, year, age, gender, marker, laboratory test used, number of individuals tested, prevalence of each marker when reported, the population group (general population, HCWs, or blood donors) and whether the data reported was for a city, sub-national (an area, region, state or province in a country) or national level, GDP per capita. In addition to HBsAg, HBeAg was recorded, as available for individuals when HBsAg was also reported. In order to record information on methodological quality and study bias resulting from non-representativeness, an additional variable was used: samples likely to be representative for the country/area specified were coded as 0 and others, e.g. convenience samples in certain communities or tribes in the country were assigned a 1, supplemented by additional information. The risk of bias/non-representativeness information was applied if the population was neither HCW nor blood donor (see description below).¹ In the following, variables extracted from the studies and assumptions made are described in detail:

1. Author, Date
2. Year start/end of study conduct: Year of study begin and end was extracted. If this information was not available from the studies, we used the commonly used assumption that the study was conducted two years prior to the year of publication (e.g. author, 2000, year of study conduct: 1998).

¹ The Newcastle Ottawa Scale for assessing the quality of nonrandomized studies in meta-analyses was consulted (http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp) and the Strobe reporting guidelines for observational studies was referred to in order to assess thoroughness of study reporting.

3. Sex: Sex-specific values were extracted. If only an overall (all) estimate was provided, the share of females in the study was specified in the column additional information.
4. Age start/end: The most specific age-group provided by the data was extracted. If the age-group on which the parameter value was based on was not available, assumptions were made based on the context of the study. Therefore, the following was applied in case of missing information on age-groups in the study population:
 - If the study was conducted in the general population without further specification and if only one prevalence estimate is provided, the age-group was considered to be 0-85 years. Subsequently, if the beginning and last age-group is missing, the lower value of the youngest age-group is 1 year, the upper value of the oldest age-groups is 85 years.
 - If the study was conducted among adult populations but no age-range is provided, the age-group is considered to be 17-65 years.
 - If the study was conducted among pupils but no age-range is provided, the age-group is considered to be 5-15 years.
 - If the study was conducted among pregnant women but no age-range is provided, the age-group is considered to be 15-49 years (reproductive age).
 - If the study was conducted among blood donors but no age-range is provided, the age-group is considered to be 17-65 years.
 - If the study was conducted among army recruits or soldiers but no age-range is provided, the age-group is considered to be 18-45 years.
 - If the study was conducted among the working population but no age-range is provided, the age-group is considered to be 16-65 years.
5. HBsAg Prevalence: The most specific prevalence estimate provided by the data was extracted (defined by age-/sex-/year-prevalence). Separate lines for each marker were used in the data extraction file (e.g. one for HBeAg and one line for HBsAg, even if the study group/publication was the same)
6. HBeAg Prevalence (optional marker): The most specific prevalence estimate (defined by age-/sex-/year-prevalence) of HBeAg among HBsAg-positive individuals was extracted and, if applicable was calculated to reflect prevalence among HBsAg carriers.

7. anti-HBc Prevalence (optional marker): The most specific prevalence estimate provided by the data was extracted (defined by age-/sex-/year-prevalence).
8. Laboratory method: Testing immune response markers of HBV infection began in the 1970s by counter-immuno-electrophoresis technique (CIEP). Since then, different detection methods have been developed (RIA, EIA, ...). The most applied method in prevalence studies is the ELISA (enzyme-linked immunosorbent assay). Five categories were established to record the method/test used for prevalence detection in the studies: ELI new (ELISA -2, -3, EIA, ...), EIA old (CMIA, CIEP, RPHA), NAT (qPCR/real-time PCR, nested PCR, multiplex PCR), other (e.g. RIA); Unknown/not specified.
9. Country: Country names were recorded according to www.who.int and, for additional analysis purpose, were grouped according to the six WHO regions: the African Region, the Region of the Americas, the Eastern Mediterranean Region, the European Region, the South East-Asia Region and the Western Pacific Region.
10. Sample size of individuals blood drawn from; of individuals involved in analyses/bases for parameter estimate: As a quality indicator of the study, we distinguished the effective sample size, i.e. the number of individuals involved in the analysis/on which the parameter estimate is based on, from the number of individuals from which blood was drawn from (separate column) and the initially calculated/planned sample size (separate column).
11. Population: Although focus was on the general population, two additional groups were included and specified. These include: HCW and blood donor (plus subgroups unspecified, paid, unpaid/voluntary). If in this column "population" was specified as HCW or blood donor and not as general population, the risk of bias column (following) remains empty.
12. Level: Information is provided if the study was conducted on a national, sub-national, city level or if the level was not further specified (four categories).
13. > Study Location: This free-text variable specifies the city/area within the country where the included study was conducted. The variables/columns Level and Study Location were additionally included following the WHO Meeting on Impact of Hepatitis B Vaccination at WHO, Geneva, in March 2014.

Additional data from other sources than the eligible studies :

1. Year of vaccine introduction in the entire country: data is derived from official reports by WHO Member States and unless otherwise stated, data is reported annually through the [WHO/UNICEF joint reporting process](http://www.who.int/entity/immunization/monitoring_surveillance/data/year_vaccine_introduction.xls?ua=1).
http://www.who.int/entity/immunization/monitoring_surveillance/data/year_vaccine_introduction.xls?ua=1
2. Period when the study was conducted: pre vaccination or post vaccination. This is determined according the year of introduction in the whole country.
3. Coverage estimates series: data is obtained from WUENIC:
http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoveragebcg.html
4. GDP per capita was used form UN data that compiles information from the World Bank Source
<http://data.un.org/Data.aspx?q=GDP&d=SNAAMA&f=grID%3a101%3bcurrID%3aUSD%3bpcFlag%3a1>),
5. Longitude and latitude data (source: www.google.com).
6. Population structure and size data for each country was from the UN population division:
<http://www.un.org/en/development/desa/population/>

6. Identify and describe any categories of input data that have potentially important biases (e.g., based on characteristics listed in item 5).

Risk of bias/representativeness: The quality of studies and data was assessed by reviewing representativeness of sampling. The risk of bias/non-representativeness column was only used if the general population (see population variable above) was focus of the study and it was distinguished between “0”= no bias/representative for the country/area/city and “1”= bias. If risk of bias/non-representativeness was indicated by the number “1”, additional information is provided in the additional information column.

Additional information: Information is provided, if risk of bias/ was indicated. Options for additional information are: either “aboriginal/tribe/indio/native” OR “high endemicity” OR “low endemicity”.

Other options here are, “xx% Females”, in case no sex-specific value is provided, “adjusted estimate”, in case no age-specific estimate is provided.

Bias factor is a dichotomous variable.

Potential important biases included geographical representation of the data points. Out of a total of 5581 HBsAg data points 851 data points were from China. Also studies were from many different sources such as blood donors and pregnant women. The former possibly having a lower proportion of Hep B prevalence than the general population as donor questionnaires often exclude individuals with risk factors for blood-borne diseases and the pregnant women possibly having a higher prevalence as were in studies to see the effect of a birth dose of vaccine to prevent vertical transmission. As the proportion of studies and size of studies that were from blood donors was significantly greater than those on pregnant women, we may presume that our estimates of prevalence of pre vaccination may be on the low side.

For data inputs that contribute to the analysis but were not synthesized as part of the study:

7. Describe and give sources for any other data inputs.

In addition to the data on HBsAg described in section 5, data inputs that were used without modification were covariates for the regression model such as GDP per capita (source: http://www.who.int/immunization/monitoring_surveillance/data/en/), longitude and latitude data (source: www.google.com). Estimated immunisation data is from the WHO-UNICEF estimates (http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoveragehepb3.html).

Additional data from other sources than the eligible studies :

1. Year of vaccine introduction in the entire country: data is derived from official reports by WHO Member States and unless otherwise stated, data is reported annually through the [WHO/UNICEF joint reporting process](http://www.who.int/entity/immunization/monitoring_surveillance/data/year_vaccine_introduction.xls?ua=1).
http://www.who.int/entity/immunization/monitoring_surveillance/data/year_vaccine_introduction.xls?ua=1
2. Period when the study was conducted: pre vaccination or post vaccination. This is determined according the year of introduction in the whole country.

3. Coverage estimates series: data is obtained from WUENIC:
http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tswucoveragebcg.html
4. GDP per capita was used from UN data that compiles information from the World Bank
 Source
<http://data.un.org/Data.aspx?q=GDP&d=SNAAMA&f=grID%3a101%3bcurrID%3aUSD%3bpcFlag%3a1>),
5. Longitude and latitude data (source: www.google.com).
6. Population structure and size data for each country was from the UN population division:
<http://www.un.org/en/development/desa/population/>

Data analysis

7. Provide a conceptual overview of the data analysis method. A diagram may be helpful.

The data was modelled using a logistic regression, weighting each study by its size and using a conditional autoregressive model accounting for spatial and economic correlations between similar countries.

See paragraph below under 8 for more details

8. Provide a detailed description of all steps of the analysis, including mathematical formulae. This description should cover, as relevant, data cleaning, data pre-processing, data adjustments and weighting of data sources, and mathematical or statistical model(s).

The data was modelled using a logistic regression, weighting each study by its size and using a conditional autoregressive model accounting for spatial and economic correlations between similar countries. The response variable in the model was the prevalence of Hepatitis surface antigen (HBsAg) with the predictor variables being age (three categories, under 5, juvenile (5-15) and adult (16+), split using the average age of participants in the study), sex (proportion female in the study), study bias (e.g. a high fraction of study participants from indigenous populations), 3 dose vaccine coverage, birth dose of the vaccine and country of study. The coverage of routine 3 dose vaccination and birth dose vaccination in each study was calculated by cross referencing the year of and age of participants in each study with the corresponding WHO-UNICEF vaccine coverage estimates for that country. The WHO-UNICEF estimates are annual data for the country as a whole, and did not contain

information on vaccine efficacy which was not used in the analysis as no data on this was obtained. The vaccine efficacy would be implicitly estimated in the analysis as we see vaccination having a variable effect across time and space across the studies.

The model was simulated in the Bayesian statistical package WinBUGS, and data manipulation and model initialisation run from R (3.3.1) using R2WinBUGS. We considered the parameters of age, sex, study bias (e.g. a high fraction of study participants from indigenous populations), vaccine coverage, birth dose of the vaccine and country of study. The coverage of routine 3 dose vaccination and birth dose vaccination was calculated by cross referencing the year of and age of participants in each study with the corresponding WHO-UNICEF vaccine coverage estimates for that country. Ages were split in to three categories, under 5, juvenile (5-15) and adult (16+).

We used the CAR-normal function, in WinBUGS, to model the spatial and developmental autocorrelation related to neighbouring countries. For each country for which we had prevalence data a weighted central position was calculated using the size and location of each study. For those with no data, we used the population centroid. In a novel approach, we considered 3 dimensions in the country proximity matrix; we used the usual geographic dimensions, latitude and longitude and also combined these with the natural log of the country's GDP per capita (see slide 1).

Country-level random effects: measures of distance

- Geographical distance (space)
 - Define neighbourhood around point of interest
 - Data points within neighbourhood contribute to estimate
 - Can either weight the points according to their distance or not
- Socio-economic distance (GDP per capita)
 - Define neighbourhood around point of interest
 - Data points within neighbourhood contribute to estimate
 - Can either weight the points according to their distance or not
- Both "distance" measures used in the model
 - Weighted 2:1 in favour of GDP distance

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Slide 1 Dimension in the country proximity matrix

This was to measure not only geographic but also the developmental proximity of countries. We normalised the geographic and GDP distance and then calculated the distance between these two

normalised figures. This creates a smoothed Gaussian surface that is dependent on both spatial proximity and GDP per-capita proximity. We compared ratios of, 1:0, 1:1, 2:1, 1:2 (Geographic:GDP). We also varied the radius of distance from which to select neighbours for the neighbourhood network, using the maximum minimum distance, twice the maximum minimum and three times the maximum minimum, thus varying the number of neighbours each country has. Finally we varied the weights of pairs of countries in the adjacency matrix, using a weighting of 1, or decaying weights over distance with $1/\text{distance}$, and $1/[\text{distance}]^2$. The outcome of these 36 different combinations led to minimum DIC (Deviance Information Criterion) being found for a ratio of 1:2 (Geographic:GDP), the neighbourhood networks minimum distance being twice the maximum minimum distance and an even weighting of $1/\text{distance}$ for each adjacent country.

The general logistic model equation is described below,

$$Y_i \sim \text{Binomial}(\pi_i, N_i), \quad \log \frac{\pi_i}{1-\pi_i} = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + u_i$$

With the spatial random effects described by

$$u_i \sim N(\bar{u}_i, \sigma_u^2/n_i)$$

where,

$$\bar{u}_i = \sum_{j \in \text{neigh}(i)} [w_{ij} u_j / n_i]$$

Where n_i is the number of neighbours for country i and weights w_i , are $1/\text{distance}$.

The model estimates the effects of each of the explanatory variables and also with the geographic and GDP levels. With these estimates it is possible to predict prevalence estimates across countries which are poorly represented in the studies. In effect, we make use of countries with many studies to inform the model and estimate the effects of the vaccination to predict the effects on countries with few. See for overview slide 2.

Estimating the HBsAg prevalence

- Logistical regression with geospatial random effects
 - Included geographic and GDP difference in a distance matrix as part of the correlated autoregressive function.
- Variables
 - sex, age, bias (e.g. indigenous peoples), vaccinated cohorts, birth dose
- $Y_i \sim \text{Binomial}(\pi_i, N_i)$, $\log \frac{\pi_i}{1-\pi_i} = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + u_i$
- Bayesian statistical package WinBUGS
- Validation on random sample of studies (10%)



Slide 2 Overview of regression model

The coverage of routine 3 dose vaccination and birth dose vaccination in each study was calculated by cross referencing the year of and age of participants in each study with the corresponding WHO-UNICEF vaccine coverage estimates for that country. The coverage of routine 3 dose vaccination and birth dose vaccination in each study was calculated by cross referencing the year of and age of participants in each study with the corresponding WHO-UNICEF vaccine coverage estimates for that country. More explicitly, we use the ages and timing of the study to calculate the years across which the participants are born, so if there was an age group range of 10-15 in a study that was undertaken in 2015, the birth years would from 2000-2005, we then average the vaccination coverage from the WHO-UNICEF estimates across those 5 years assuming that each age was evenly represented in that age group in the study. The same process was used for the 3 dose and birth dose vaccination.

We estimated prevalence pre and post vaccination by using the appropriate levels of each fixed effect calculated from the model. So in the no vaccination scenario, vaccine and birth dose effect were scaled to zero. In the 2015 scenario with vaccination, the relevant age groups received the appropriate level of vaccination (WHO-UNICEF estimates) of both the 3-Dose vaccination and the birth dose vaccination to estimate the levels of prevalence. The overall average across each age category was informed and weighted by the population structure and size (UN population data).

9. Describe how candidate models were evaluated and how the final model(s) were selected.

See paragraph below under 13

12. Provide the results of an evaluation of model performance, if done, as well as the results of any relevant sensitivity analysis.

See paragraph below under 13

13. Describe methods for calculating uncertainty of the estimates. State which sources of uncertainty were, and were not, accounted for in the uncertainty analysis.

This model structure produces estimates for all fixed effects and also individual country level risk, this provides information on which are significantly at greater or lower risk to the average risk.

All parameters were given un-informative priors. Simulations were run with 3 MCMC chains with 50,000 burn in iterations and each parameter estimated from 1000 samples taken from a thinned 250,000 iterations to produce the posterior distribution. Convergence was attained, with \hat{r} values all very close to 1.000. Due to the Bayesian framework and WinBUGS software it was possible to gain estimates for countries where we had no data on prevalence, using their GDP and geographic proximity to inform this estimate. Those countries with the largest number of studies provided the estimates with the tightest confidence intervals and those with few or no data were less well defined, often producing a log normal distributed posterior distribution, giving estimates with long tails.

Posterior distributions of parameters were inspected for convergence and to check for covariance between parameters. Where necessary parameters were centred and scaled to $N(0, 1)$ to aid parameter convergence and remove covariance. This was done for the sex parameter, which was entered as the proportion of the sample that was female; this was seen to co-vary with the intercept and bias parameters before re-centring and scaling. However, the covariance of routine vaccination and birth dose persisted even after re-centring. This is in part unsurprising as there are a few instances where birth dose is administered without the routine vaccination. Here we tried to reduce this interaction of the terms by transforming the birth dose data. We modelled birth dose using only data where the birth dose was greater than 60, 70, 80 & 90% respectively, we also modelled birth dose to the square, thus increasing the effect of high birth doses over smaller doses. Model selection dependent on which one both reduced the covariance between the parameters and returned the lowest DIC score.

Model validation was conducted using 90% of randomly selected data against the remaining 10%, and by comparing model estimates of prevalence against observed data (Figure 2). Figure 3 shows the average prevalence in each country from all the studies plotted against the models estimate. Figure 4 shows the marginal and joint posterior distributions for the fitted parameters. Table 1 gives the estimated parameter values with associated credible intervals.

During the validation exercise (in which countries were consulted over their estimates) it was pointed out that China had undertaken three very large-scale population-based serological surveys in order to establish baseline prevalence and progress towards HBV elimination. There were a large number of other surveys from China, that are less representative than these three nationwide surveys. We conducted a sensitivity analysis by restricting the data from China to the three nationally representative surveys. The effect of this change in input data was that the effect of vaccination was more distinct, but the estimated age effects (change in prevalence in children under 5, or juveniles (children 5-15 years)) were no longer significantly different from zero (see Table 2 and Figure 5). The deviance was significantly reduced, suggesting a much better fitting model (Table 2), albeit on a somewhat reduced dataset.

10. State how analytic or statistical source code used to generate estimates can be accessed.

See under Section 13 The code will be attached to the published manuscript.

Results and Discussion

11. Provide published estimates in a file format from which data can be efficiently extracted.

We estimated the prevalence of chronic HBV infection (HBsAg seroprevalence) based on data from all eligible studies for each country. 95% credible intervals (CIs) were obtained from the model.

For regional prevalence estimates, countries were grouped into six WHO regions: Africa, the Americas, the Eastern Mediterranean, Europe, South East Asia, and the Western Pacific. Regional HBsAg prevalence estimates were produced from country-specific estimates, weighted according to the population size of each country. Similarly, regional 95% CIs were obtained from weighted country-specific variances of the prevalence estimate from the model.

Country and regional level results including the confidence intervals can be found in enclosed MS Excel file.

12. Report a quantitative measure of the uncertainty of the estimates (e.g. uncertainty intervals).

We estimated the mean prevalence for each country from the mean model results using the inverse logit. 95% credible intervals on the prevalence were generated from the appropriate bounds on the model results in a similar way. Regional estimates were obtained by summing over each country in the region.

17. Interpret results in light of existing evidence. If updating a previous set of estimates, describe the reasons for changes in estimates.

Concerns were expressed about the level of analysis of the very large data set used by the existing study from Schweitzer et al (2015). In particular how age was dealt with; how vaccination effects by time period influenced the results and in general that there was much more information than was expressed by the published paper.

It was noted that the extracted data was all available and that the London School of Hygiene and Tropical Medicine (LSHTM) were adding exact geographical locations to each paper.

The point was made that sub-national analyses are of increasing interest, that analyses of women of child bearing age are of particular importance and that some form of age stratified estimates (particularly children under five and adults) was needed.

It was noted that voluntary blood donor data was included in the database which is likely to be low risk and should be analysed separately. It was also noted that migrants, high risk populations and indigenous populations were excluded and therefore the estimates need to be treated interpreted carefully.

It was emphasised that the database should be updated regularly – preferably annually. It was also noted that the specific evaluations of vaccination programmes should be collected as a sub-set of the data to allow future analyses of them separately.

The clean dataset will be available to the public domain in the Virtual Hub on WHO Vaccine Preventable Diseases

18. Discuss limitations of the estimates. Include a discussion of any modelling assumptions or data limitations that affect interpretation of the estimates.

Similar to the Schweitzer (2015) our burden estimates exclude studies specific to population groups at particular risk of chronic HBV infection, such as migrants from endemic areas resident in low-prevalence countries. Since in some countries these groups form a substantial part of the population and represent the majority of those living with chronic HBV infection this exclusion leads to an underestimate of the total burden of HBV.

Table 1. Parameter estimates and 95% credible interval from logistic regression for a) base-line model, and b) model including only the nationally representative data from China.

a)

Parameters	Mean	2.50%	97.50%
Intercept	-3.57	-3.69	-3.44
vac.effect	-0.88	-1.53	-0.30
BD.effect	-0.79	-1.37	-0.21

juv.effect	-0.11	-0.24	0.03
bias.effect	0.56	0.38	0.75
female.effect	0.10	0.00	0.22
child.effect	-0.11	-0.30	0.08
deviance	528369	509800	592483

b)

Parameters	mean	2.50%	97.50%
intercept	-3.61	-3.75	-3.47
vac.effect	-1.36	-1.71	-1.05
BD.effect	-1.10	-1.52	-0.66
juv.effect	-0.08	-0.22	0.06
bias.effect	0.92	0.66	1.16
female.effect	0.24	-0.13	0.63
child.effect	-0.01	-0.17	0.16
deviance	325,306	309,400	381,790

Figure 1 PRISMA flowchart diagram

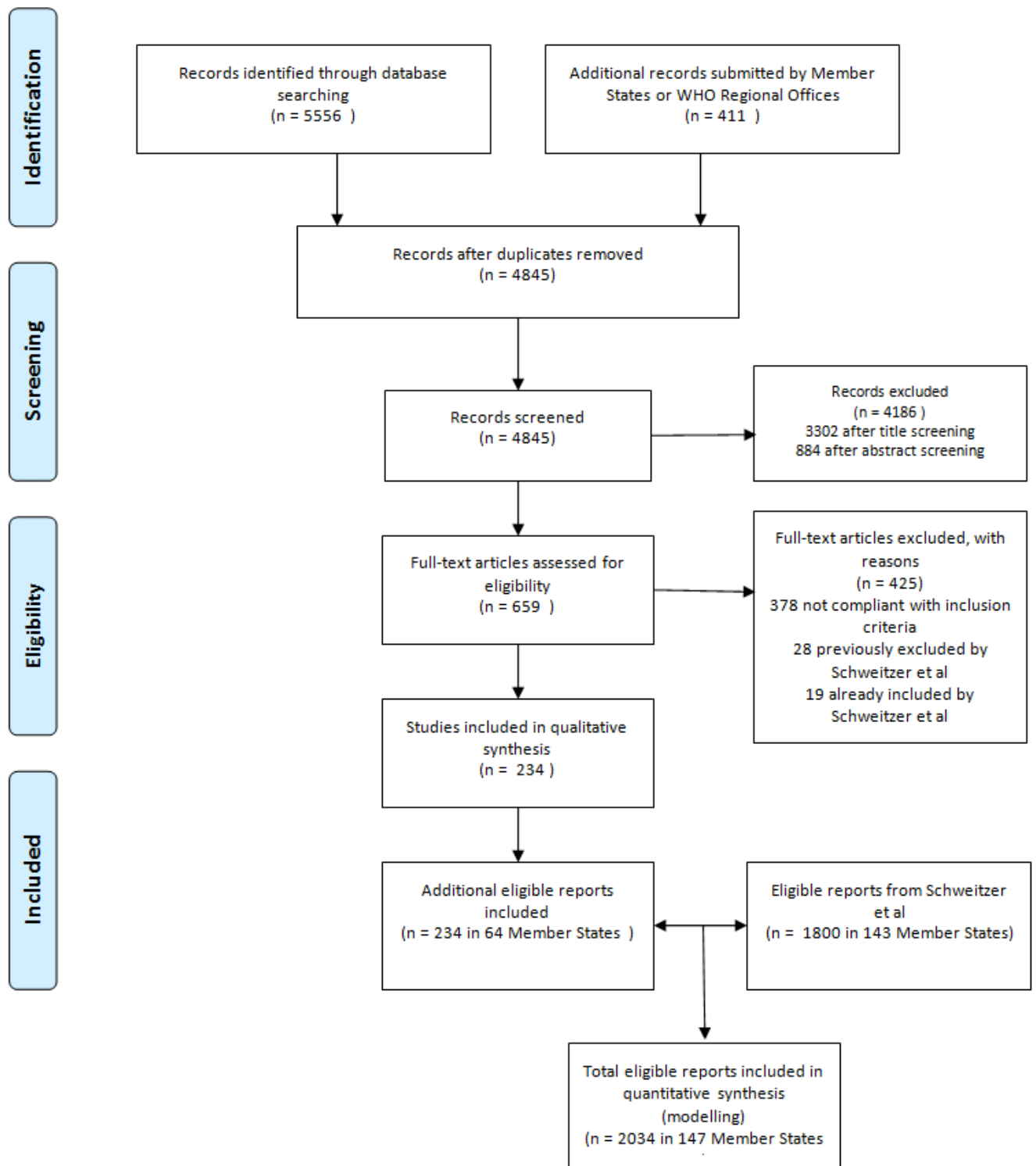


Figure 2. Predictions of 10% of study estimates using 90% of randomly selected studies to inform the model. 95% C.I. are shown in each axes dimension.

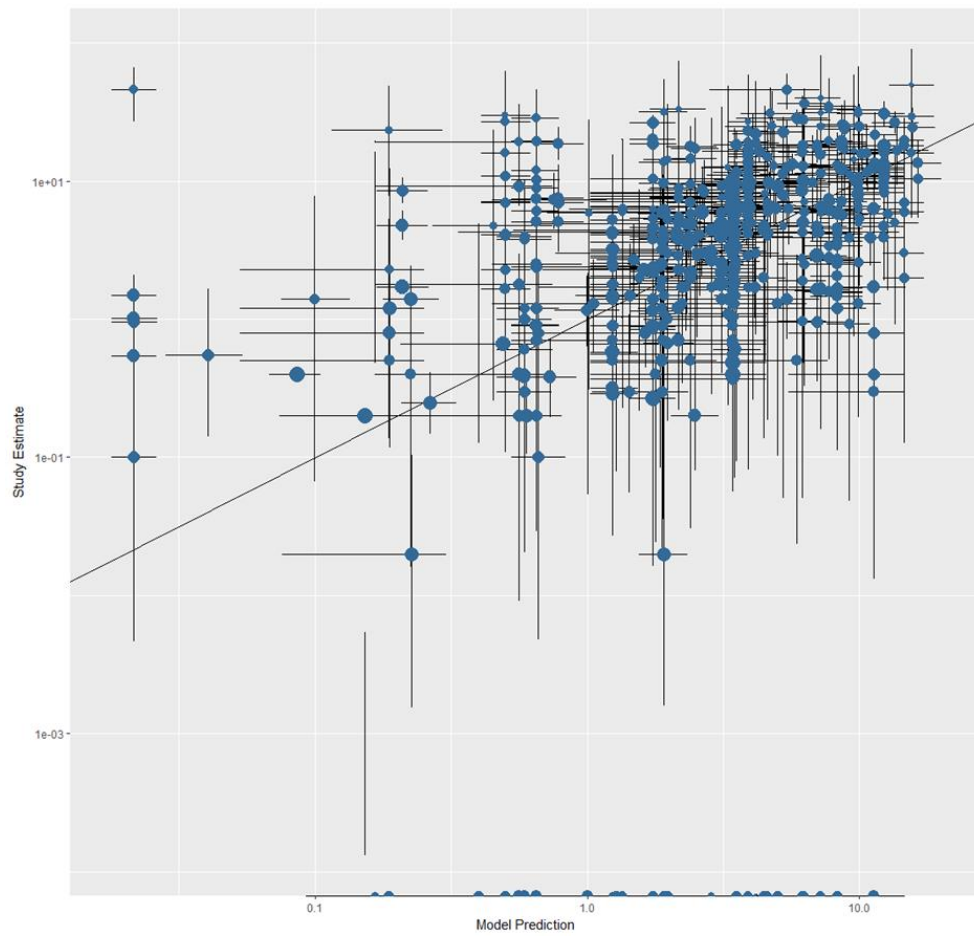
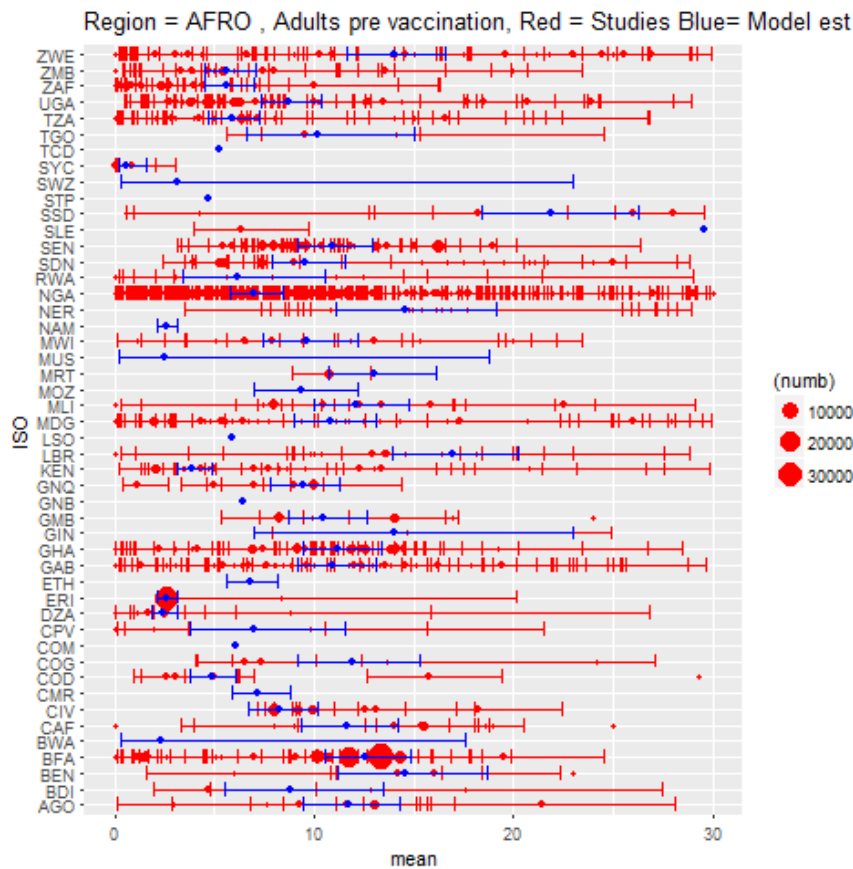
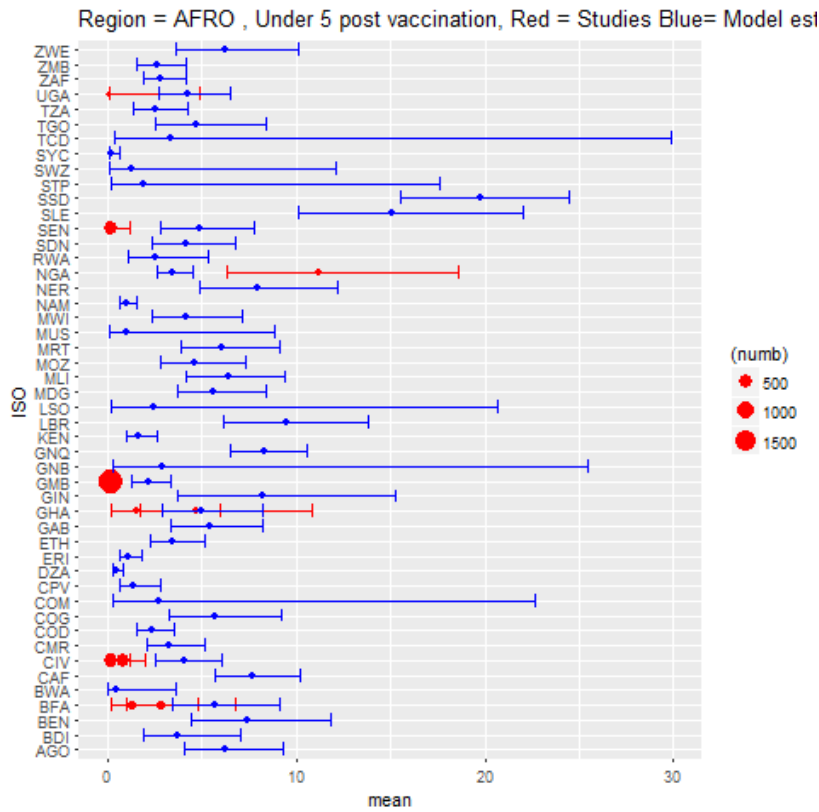
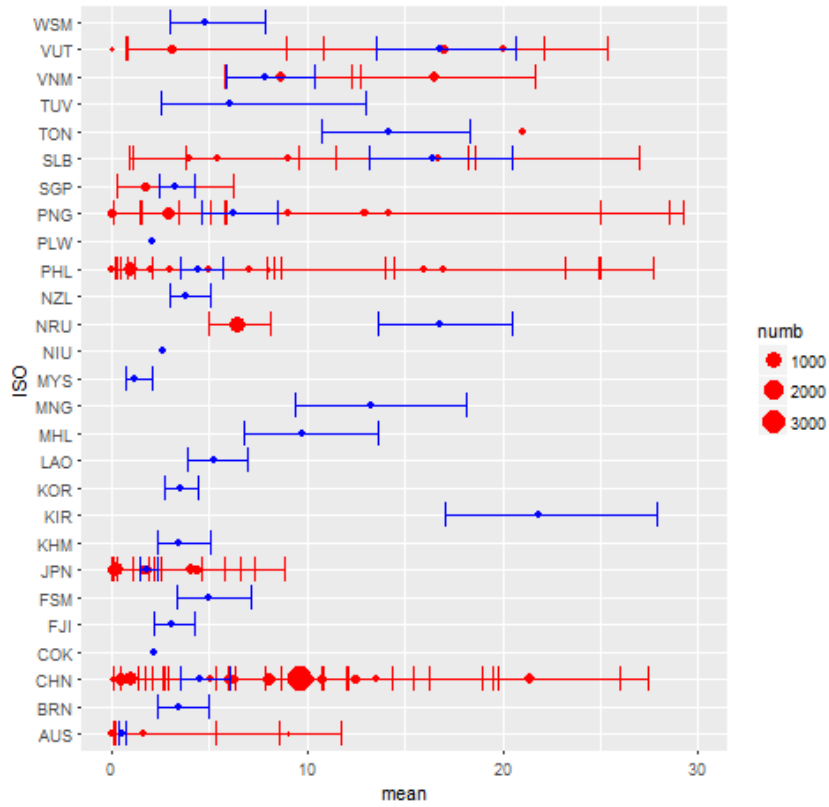


Figure 3. Comparison of observed data (red points and associated 95% confidence intervals, the size of the point represents the sample size of the survey) with model estimates of prevalence and 95% credible intervals (blue dots and points) by country, for three groups: pre-vaccination children under 5, post-vaccination children and adults (prevaccination). Data and estimates for 2 regions are shown, AFRO (first 3 plots) and WPRO (second 3 plots).

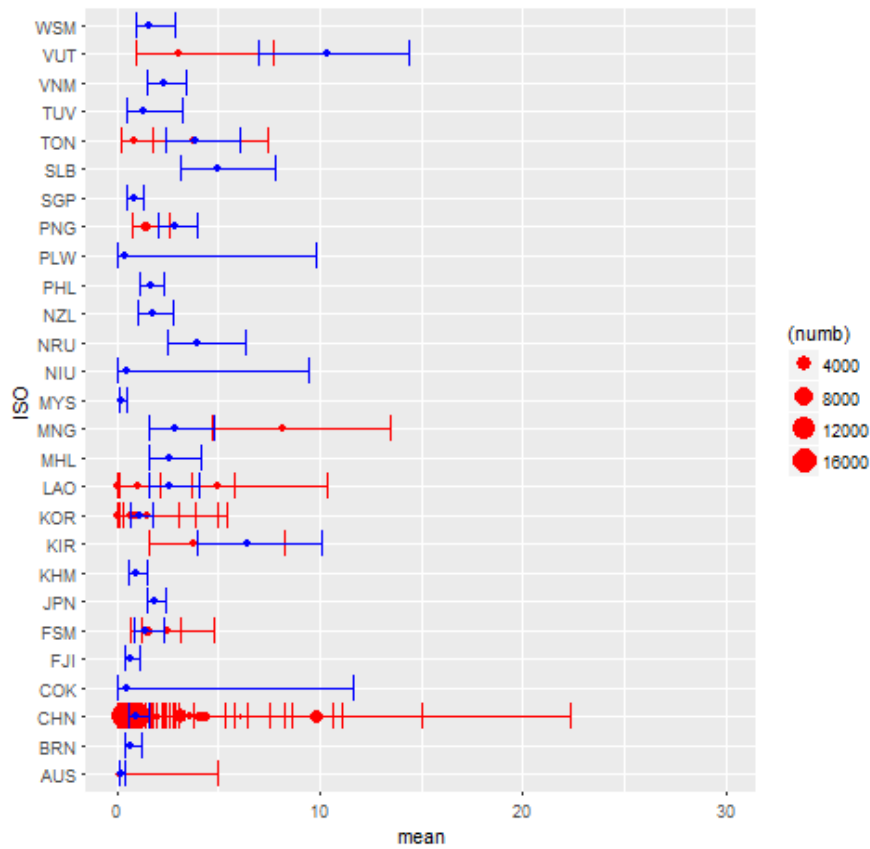




Region = WPRO , Under 5 pre vaccination, Red = Studies Blue= Model est



Region = WPRO , Under 5 post vaccination, Red = Studies Blue= Model est



Region = WPRO , Adults pre vaccination, Red = Studies Blue= Model est

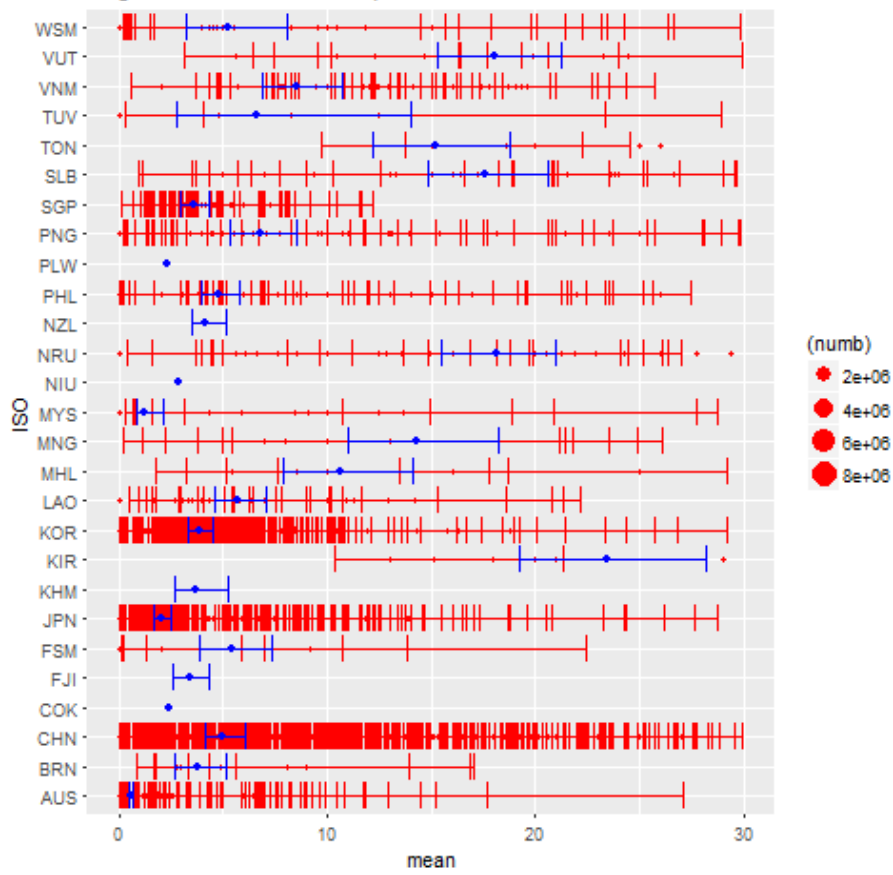
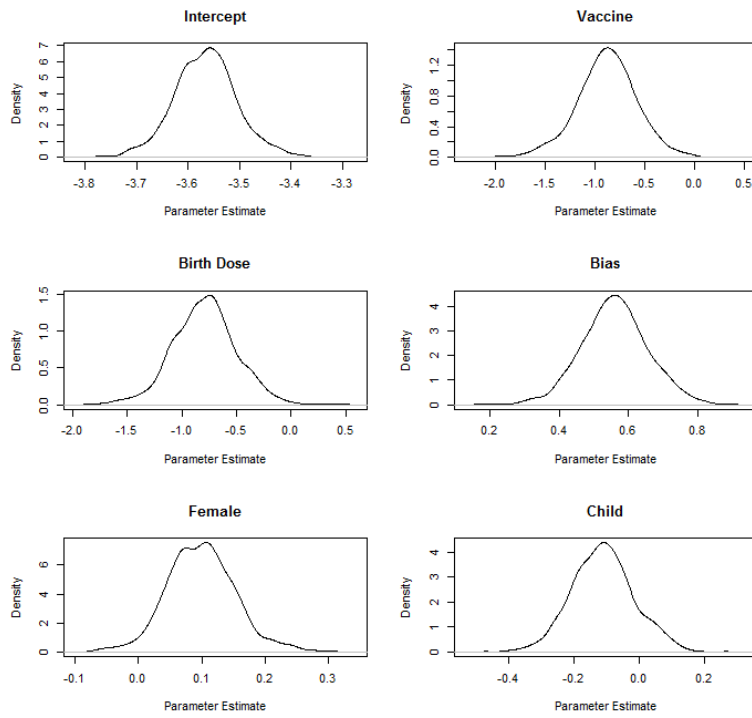


Figure 4. Marginal posterior distributions for parameters from the base-case model (a) and correlations between parameters for this model (b).



b)

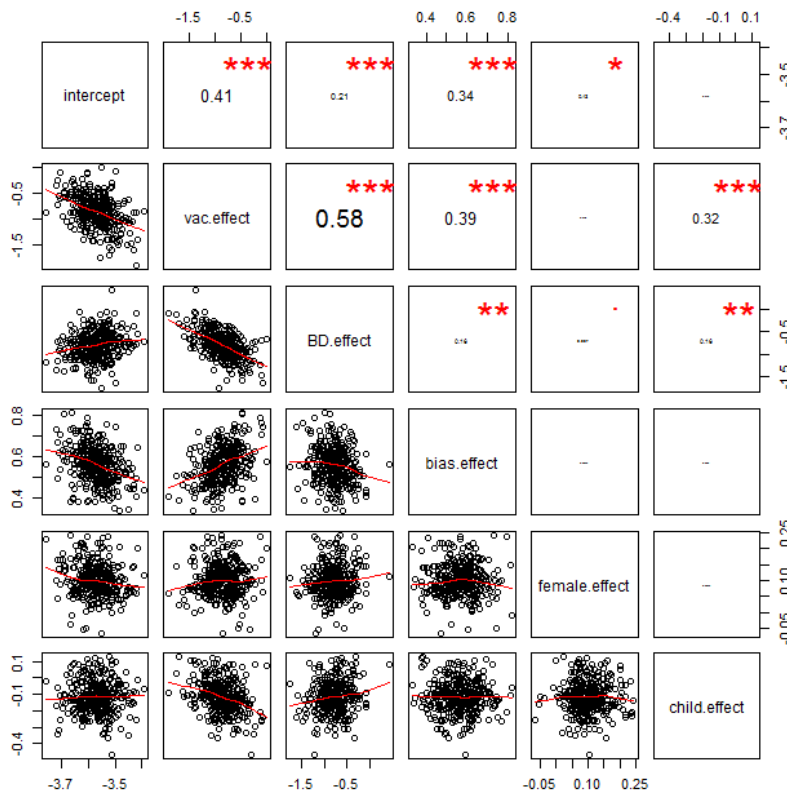
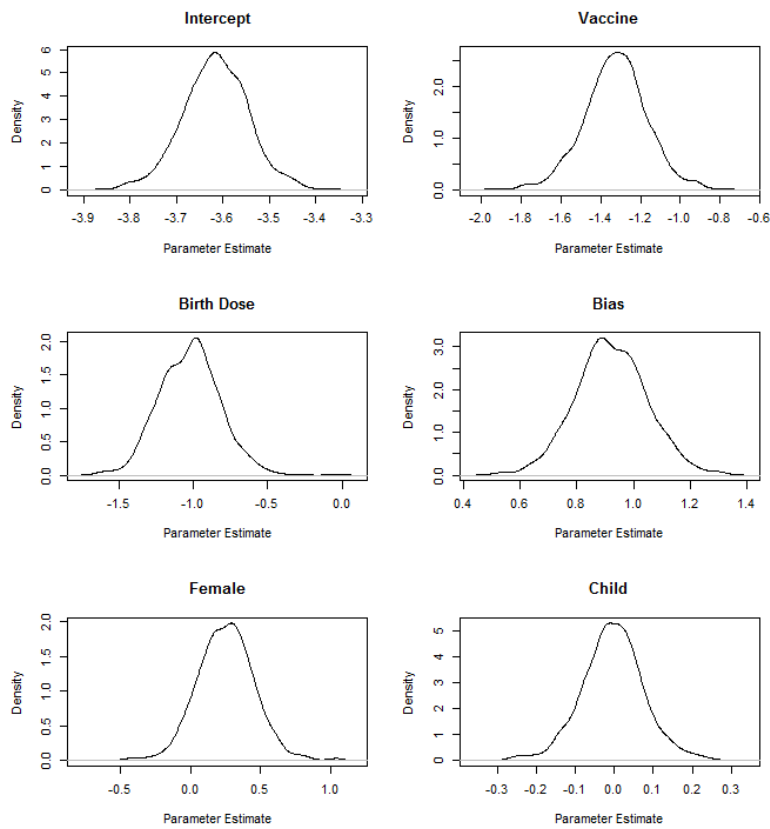


Figure 5. Marginal posterior distributions for parameters from the model excluding non-representative data from China (a) and correlations between parameters for this model (b).

a)



b)

